Write Channel Model for Bit-Patterned Media Recording

Aravind R. Iyengar, Paul H. Siegel, *Fellow, IEEE* and Jack K. Wolf, *Life Fellow, IEEE*

Abstract—We propose a new write channel model for bit-patterned media recording that reflects the data dependence of write synchronization errors. It is shown that this model accommodates both substitution-like errors and insertion-deletion errors whose statistics are determined by an underlying channel state process. We study information theoretic properties of the write channel model, including the capacity, symmetric information rate, Markov-1 rate and the zero-error capacity.

Index Terms—Bit-patterned media, High-density magnetic recording, Channel capacity, Symmetric information rate, Markov-1 rate, Zero-error capacity.

I. INTRODUCTION

MAGNETIC recording channels are typically modeled as binary-input intersymbol interference (ISI) channels, also called *partial response* channels [2], [3]. An implicit assumption made in these channel models is that the data are correctly written on the disk and that errors occur only during the readback process. While this is a realistic assumption in conventional recording on continuous media, it is questionable in the context of certain advanced recording technologies, such as bit-patterned media (BPM) recording, that may be required to achieve higher storage densities. In this paper, we will examine some of the underlying causes of errors in the recording process, particularly in BPM recording, and then propose a new probabilistic write channel model that captures some of the data dependence of these write errors. Thus, the input to this write channel model is the data sequence to be recorded and the channel output is the "noisy" sequence that actually gets stored on the medium. This leads to a description of the full data recording and readback process as a cascade of an imperfect write channel and a noisy (partial response) readback channel.

Hu et al. [4] proposed a model for the BPM recording channel in which the write channel was a binary symmetric channel (BSC), and the readback channel was a linear, intersymbol-interference (ISI) channel with additive noise. They proposed and evaluated detection methods for this channel, and they investigate achievable information rates for such a system. In [5], we considered an idealized cascaded channel model in which the write channel was again a BSC and the readback channel was a memoryless, binary-input, additive white Gaussian noise (AWGN) channel. We studied theoretical properties of this channel, as well as the decodable regions of LDPC codes under a number of decoding algorthms. In this paper, we focus on a new write channel model, and we determine and compare bounds on several relevant information-theoretic limits: capacity, symmetric information rate (SIR), information rate with first-order Markov inputs (Markov-1 rate), and zero-error capacity.

The remainder of the paper is organized as follows. In Section II, we start with a brief description of the write process in bit-patterned media recording, highlighting the main factors leading to write errors. We review some useful mathematical notation in Section III-A, and then present the new probabilistic, data-dependent write channel model in Section III-B. We discuss the types of write errors that can occur – including substitution-like errors and insertion-deletion errors – and explain their connection to the channel state process that underlies the model. Two classes of channel state processes are introduced: a Bernoulli state process (Section III-B1) and a binary Markov state process (Section III-B2). In Section III-C we make the observation that, although proposed as a model for BPM recording, the Markov state channel model is also relevant to high-density magnetic recording using conventional granular media. In Section IV we give bounds on the capacity, the SIR, and the Markov-1 rate of the Bernoulli state channel. Section V gives similar bounds on the capacity for the binary Markov state channel. The SIR is numerically computed for both of the channels considered in Section III. In Section VI, we introduce a generalization of binary Markov state channels, namely the K-ary Markov state channels. For one such generalized channel, we numerically estimate the SIR and derive bounds on the channel capacity. Finally, in Section VII, we explore the zero-error capacity of the proposed class of write channel models. Section VIII provides concluding remarks.

II. BIT-PATTERNED MEDIA RECORDING

A conventional magnetic recording medium is a continuous film of magnetic *grains* that coats the surface of the disk substrate. Each grain is an atomic magnetic unit that assumes one of two possible magnetic states. A group of grains together form a *bit-cell*, an entity that stores one bit of information. Therefore, as the areal information density is increased, the number of grains forming a bit-cell reduces. One of the problems with high-density magnetic recording in conventional

media is the *super-paramagnetic effect*, wherein the magnetic states of individual grains change due to the influence of neighbouring grains or due to changes in temperature. When the areal information density is increased to point where there are only a few grains per bit, such uncontrolled changes in the magnetic states of grains are detrimental to reliable information storage.

Bit-patterned media recording (BPMR) proposes to get around this problem by making use of patterned magnetic islands separated by non-magnetic material [6]. However, this new structure of the magnetic medium introduces technical challenges not seen in recording on conventional media. An immediate requirement of this media structure is near-perfect synchronization of the write process to ensure that the write head is positioned correctly over the magnetic islands on this disk, i.e., to ensure that the head is positioned within the so-called *writing window zone* of the islands [7]. Assuming that this write synchronization is achieved through the use of timing synchronization patterns, there is a possiblity of frequency and/or phase mismatch leading to incorrect writing. Furthermore, even without a timing mismatch, imperfections in the configuration of the patterned magnetic islands may cause writing errors. Finally, as in conventional magnetic recording, the *switching field distribution* of magnetic grains may contribute to errors in the write process. We will refer to write errors induced by any of these mechanisms as *written-in errors*.

Another important feature of BPMR is the geometry of the writing process. Along the *down-track* direction, the span of influence of the magnetic write field is typically larger than the spacing between the islands. Therefore, at any given time, the write head influences multiple adjacent islands. So, in the process of recording a bit on a specified island, the bit value is also recorded on a certain number of subsequent islands, with these islands themselves being overwritten in the future by subsequent bits. Fig. 1 gives an illustration of this idealized write process. We refer to the number of subsequent islands influenced by the write head as the *writing span*, D. We will assume throughout this paper that $D \geq 1$, which implies that write process has memory and, as a consequence, the write channel is data-dependent.

III. WRITE CHANNEL MODEL

A. Notation

We will denote random variables using capital letters X, Y, Z; random vectors as $X_i^j = (X_i, X_{i+1}, \cdots, X_j), i, j \in \mathbb{N}$ when $i \leq j$, a null variable when $i > j$ or when j is non-positive. Infinite-dimensional random vectors are considered to be discrete-time random processes and are denoted using calligraphic letters $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$. The alphabets over which random processes are defined are denoted by $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{Z}$. Unless otherwise stated, we will denote the channel input variable, process and alphabet using the letters X, $\mathcal{X}$ and $\mathbb{X}$ respectively. Similarly, Y, $\mathcal{Y}$, $\mathbb{Y}$ will correspond to the channel output, and Z, $\mathcal{Z}$, $\mathbb{Z}$ to the channel state ($\mathbb{Z}$ is not to be confused with the set of integers $\mathbb{Z}$). For a real number $a \in [0, 1]$, we define $\overline{a} = 1 - a$.
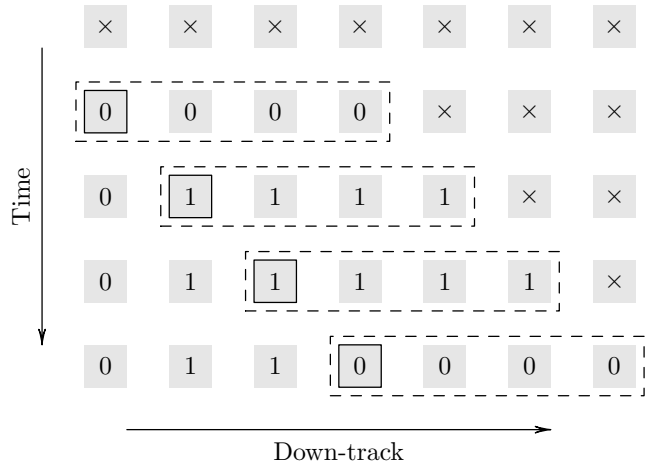

Fig. 1. Example illustrating writing one bit on each island in BPMR with a write head of writing span $D = 3$. Depicted here are snapshots of the top view of the disk and write mechanism. Initially, all islands, represented here using gray squares, have unknown magnetic states represented by $\times$, which could be 0 or 1. At the first time instant, the write head, shown here using a dashed rectangle, is positioned over the first four islands. In this position, the first island is written with the data bit 0. The three subsequent islands are also written with the first data bit 0, before being overwritten with their own data. After each bit is written, the head is moved over to the next island along the down-track direction. At each position of the write head, the corresponding "current" island is highlighted with a black square boundary.

B. Model

We assume that the input, output and the channel state alphabet is the binary set $\mathbb{X} = \mathbb{Y} = \mathbb{Z} = \{0, 1\}$. The write channel model can be written as

$$Y_i = X_i \oplus (X_i \oplus X_{i-1}) \otimes Z_i, \tag{1}$$

where the state process $\mathcal{Z}$ is independent of the inputs and the outputs, and $\oplus$ and $\otimes$ represent addition and multiplication in $GF(2)$, respectively. The channel state Z_i is not to be confused with the magnetic state of the island.

It is natural to assume that the channel input and output alphabets are binary since the (intended and actual) magnetic states of the islands can be one of two possible states. The channel state Z_i is a random variable that represents a failure in writing due to one of the conditions mentioned in the previous section, i.e., when $Z_i = 1$ the write head can be assumed to have failed in writing the intended bit. However, this does not necessarily imply a written-in error because there would be no error if the bit to be written is the same as the existing magnetic state of the island. This is captured by the term $(X_i \oplus X_{i-1})$ that is multiplied with Z_i in (1). This "noise" term $(X_i \oplus X_{i-1}) \otimes Z_i$ is justified because we assume that the timing mismatch or the irregularity of island patterns can cause the write head to be positioned, in the worst case, on the island immediately following the correct island. We will continue with this assumption until Section VI, where we construct a more general write channel model. Also note that when the first bit is written late, i.e. when $Z_1 = 1$, we have $Y_1 = X_0$ which represents the pre-existing magnetic state of the first island. We will assume that X_0 is equally likely to be a 0 or a 1. Similarly, when the last bit is written late, the last island has a bit in error if the last and the penultimate bits are

different. In either case, n magnetic islands are read and their contents are interpreted as the n data bits so that there is no blocklength mismatch.

Based on the Z_i sequence, the channel in (1) appears to produce different types of errors. When the Z_i sequence consists of isolated ones, the channel appears to produce *substitution* errors. This is illustrated in the Fig. 2. Such

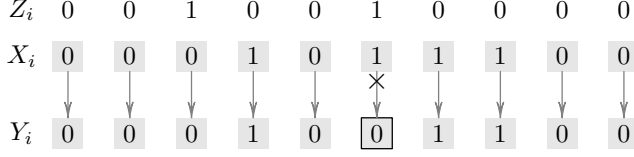| Z_i	0	0	1	0	0	1	0	0	0	0
X_i	0	0	0	1	0	1	1	1	0	0
Y_i	0	0	0	1	0	0	1	1	0	0

Fig. 2. An example of substitution-like errors produced by the write channel. The sequence Z_i gives the channel state for each magnetic island. In this example, $Z_i = 1$ for the third and the sixth islands. The sequence X_i gives the intended magnetization of the islands, i.e., the data to be written. Taking into account the channel states Z_i, the sequence Y_i shows the resulting island magnetizations. Note that substitution-like errors occur only when $Z_i = 1$, as was the case with the sixth island here, highlighted with a box around the island in the Y_i sequence. However, not every $Z_i = 1$ results in an error, as illustrated by the third island.

substitution-like errors can occur when the write head misses islands at random and independently of its success in writing on previous islands. Noting that when $Z_i = 0$, $Y_i = X_i$ so that the output reproduces the input exactly, and when $Z_i = 1$, $Y_i = X_{i-1}$ so that the output reproduces the input with a delay of one time instant, we can see that another way to write the relation in (1) is

$$Y_i = X_{i-Z_i}. \qquad (2)$$

Now, when the Z_i sequence consists of long runs of ones, the channel appears to produce paired *insertion-deletion* errors, with insertions accompanying $0 \to 1$ channel state transitions and deletions accompanying $1 \to 0$ channel state transitions, as shown in Fig. 3. These insertion-deletion errors can happen

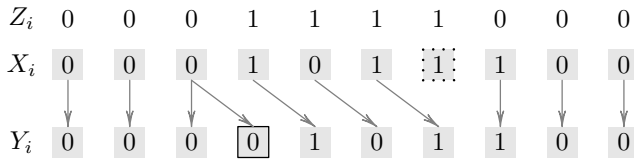| Z_i	0	0	0	1	1	1	1	0	0	0
X_i	0	0	0	1	0	1	1	1	0	0
Y_i	0	0	0	0	1	0	1	1	0	0

Fig. 3. An example of an insertion-deletion error. Due to the channel state sequence Z_i, the channel inputs are transformed as shown by the arrows between the X_i and Y_i sequences with the resulting magnetic states of the islands shown in the sequence Y_i. The inserted bit in the Y_i sequence accompanying the channel state transition $0 \to 1$ is shown with a box around the corresponding island. The deleted bit from the X_i sequence accompanying the channel state transition $1 \to 0$ is shown with a dotted box around the island.

as a result of timing synchronization errors, wherein there is a frequency mismatch between the islands and the write head; or as a result of a group of islands being separated farther than usual or having larger switching fields.

The channel in (1) and (2) is completely defined by specifying the statistics of the $\mathcal{Z}$ process. We will consider the channel under two different statistical assumptions on the $\mathcal{Z}$ process in the following, and show how this difference in statistics changes the typical behaviour of the channel.

1) Bernoulli state channel: When the channel state process $\mathcal{Z}$ is an i.i.d. Bernoulli(p) process, i.e., $\mathcal{Z} \sim \mathcal{B}(\mathsf{p})$, we will call the channel the *Bernoulli state channel*. In this case, the channel is completely specified by the parameter p. We shall henceforth denote this channel by $W(\mathsf{p})$. Typically, for small values of the parameter p we can expect this channel to produce errors resembling substitution errors (See Fig. 2).

2) Binary Markov state channel: When the channel state process $\mathcal{Z}$ is a first-order binary Markov process[1] $\mathcal{Z} \sim \mathcal{M}_1^{(2)}(\mathsf{p_i}, \mathsf{p_d})$, i.e., $\mathbb{P}\{Z_i | Z_1^{i-1}\} = \mathbb{P}\{Z_i | Z_{i-1}\}$, $\mathbb{P}\{Z_i = 1 | Z_{i-1} = 0\} = \mathsf{p_i}$ and $\mathbb{P}\{Z_i = 0 | Z_{i-1} = 1\} = \mathsf{p_d}$, we will refer to the channel as the *binary Markov state channel* and denote it by $W(\mathsf{p_i}, \mathsf{p_d})$. As noted earlier, we can expect such a channel to typically produce paired insertion-deletion errors (See Fig. 3). Hence, the parameters $\mathsf{p_i}$ and $\mathsf{p_d}$ can be thought of as insertion and deletion probabilities, respectively, of the channel.

We note here that this channel differs from the most general insertion-deletion channel in two ways. First, the inserted bit is always the same as the last written bit. Whereas this seems to be a serious limitation of the model in comparison with the general insertion channel model, it is to be noted that in most practical insertion-deletion channels—channels with synchronization errors, or sticky channels [8]—the inserted bit is usually the one last written (or transmitted), rather than being a random bit. Second, the insertions and deletions are paired so that one never sees two consecutive insertions or deletions. While this seems to limit the scope of the channel, a straightforward extension to channels with a finite maximum number of consecutive insertions or deletions is possible. However, the estimation of the channel characteristics becomes more complex as the maximum number of consecutive insertions increases (See Section VI). Nevertheless, the binary Markov state channel gives us good insight into more general insertion-deletion channels. Moreover, unlike much of the previous work in the literature wherein channels cause only insertion [8] errors or only deletion [9]–[14] errors, this model considers both insertions and deletions in the same setting. Among works that handle both insertions and deletions, the set up in this paper resembles the duplication-deletion channel of [15] more than it does the generic insertion-deletion channel of [16]. However, the channel model considered here differs from those in the aforementioned papers in that the insertion-deletion process has memory, i.e. the insertions and deletions are not i.i.d.

C. High-density recording with granular media

The channel model in (1) can also be used to describe high-density magnetic recording on conventional granular magnetic media [17]. Although media granularity is typically considered as a two-dimensional phenomenon, we consider granularity only in one-dimension – along the down-track direction, and assume adjacent tracks to be independent. This simplification

[1]The superscript in the notation $\mathcal{M}_1^{(2)}(\mathsf{p_i}, \mathsf{p_d})$ denotes the alphabet size over which the process is defined and the subscript denotes the memory in the process. The two arguments give the transition probabilities between the two states.

allows us to establish lower bounds for performance over the two-dimensional channel as proposed in [17].

Media granularity in one-dimension results in written-in errors as follows. At storage densities of the order of 1 bit per grain, the variation in grain size plays an important role in deciding the reliability of data storage. In this case, bit-cells are at most as large as individual grains. We assume that the grains are all 1 or 2 bit-cells in size, and that the magnetic state of each grain is decided by the last bit written on them. This is depicted in Fig. 4. Grains that are 2 bit-cells large will
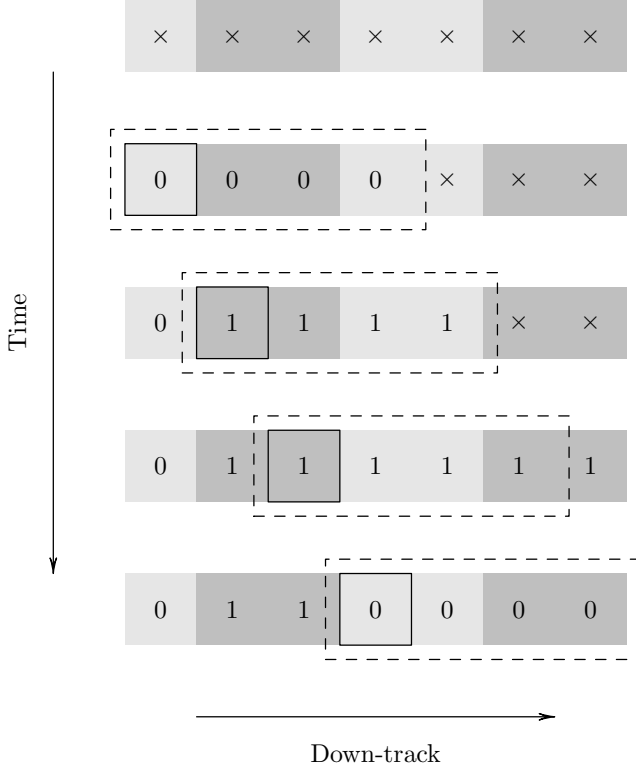

Fig. 4. Writing on granular media. Individual bit-cells are shown as ×'s in the first row, corresponding to the magnetic states read from these cells, which could be 0's or 1's. Also shown in shades of gray are the magnetic grains: grains that comprise one bit-cell are shown in light gray and those that comprise two bit-cells are in a darker shade of gray. The two bit-cells comprising a grain of size 2 always have the same magnetic state. As in the case of BPMR, the write head spans multiple bit-cells, as shown by the dashed rectangles. The "current" bit-cell at each time instant is shown with a black square.

result in written-in errors if the two bits written on them are different. Fig. 5 gives an illustration of the written-in errors in this scenario. Comparing Figures 3 and 5, it is easy to see that the channel model in this case can be written as

$$Y_i = X_i \oplus (X_i \oplus X_{i+1}) \otimes Z_i.$$

Using a simple time-reversal argument, it can be seen that this is exactly the channel in (1). However, in this case, the Z_i sequence cannot have two consecutive ones, i.e., it satisfies the $(1, \infty)$ run-length constraint [18, Chap. 4]. We choose to model this as the $W(\mathsf{p}_i, 1)$ channel since any realization of the $\mathcal{M}_1^{(2)}(\mathsf{p}_i, 1)$ process meets the run-length requirement. In [19], the authors consider this channel and derive upper bounds for the achievable rates over the channel. Our focus here will be
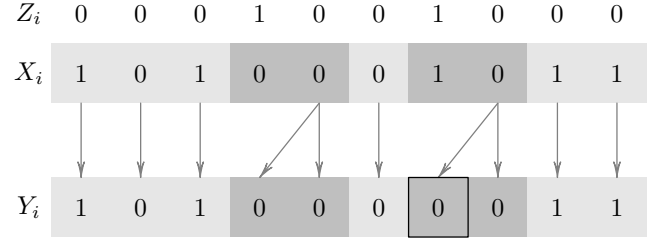

Fig. 5. Written-in errors due to media granularity. The $\mathcal{Z}$ process represents the grain pattern with Z_i being 1 when the corresponding bit-cell is the first bit-cell in a grain that comprises two bit-cells. The X_i sequence shows the data to be written. The arrows between the X_i and Y_i sequences show the transformation of information according to the grain pattern, resulting in the sequence Y_i being stored. The written-in errors are shown with a box around the corresponding bit-cell.

on obtaining bounds on the achievable information rates for the general channel in (1) in the context of the two channel state processes defined in Section III-B.

IV. BERNOULLI STATE CHANNEL

The channel space for the Bernoulli state channel defined in the previous section is parameterized by $\mathsf{p} \in [0, 1]$. Since the channel has memory, its capacity is given as

$$C = \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I\left(X_1^n; Y_1^n\right).$$

All the entropy and mutual information terms calculated are in bits, so that the capacity is always measured in bits per channel use. We will denote by $C(\mathsf{p})$ the capacity of the Bernoulli state channel with parameter p.

Proposition 1 (Channel symmetry): $C(\mathsf{p}) = C(\overline{\mathsf{p}})$.

Proof: Let us denote by $I_{\mathsf{p}}(X_1^n; Y_1^n)$ the mutual information between the vectors X_1^n and Y_1^n when the channel parameter is p. Note that the channel output can be simultaneously written as

$$Y_i = (X_i \otimes \overline{Z}_i) \oplus (X_{i-1} \otimes Z_i)$$

and

$$Y_i = (S_i \otimes Z_i) \oplus (S_{i+1} \otimes \overline{Z}_i)$$

where $S_i = X_{i-1}$ and $\overline{Z}_i = Z_i \oplus 1$. Thus

$$I_{\mathsf{p}}(X_i; Y_1^n | X_1^{i-1}) = I_{\overline{\mathsf{p}}}(S_{n-i+1}; Y_1^n | S_{n-i+2}^n)$$

for all but a vanishing fraction of indices $i \in \mathbb{N}$, as $n \to \infty$. Therefore,

$$C(\mathsf{p}) = \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{\mathsf{p}}(X_1^n; Y_1^n)$$

$$= \lim_{n\to\infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} \sum_{i=1}^{n} I_{\mathsf{p}}(X_i; Y_1^n | X_1^{i-1})$$

$$= \lim_{n\to\infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} \sum_{i=1}^{n} I_{\overline{\mathsf{p}}}(S_{n-i+1}; Y_1^n | S_{n-i+2}^n)$$

$$= \lim_{n\to\infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{\overline{\mathsf{p}}}(S_1^n; Y_1^n)$$

$$= \lim_{n\to\infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{\overline{\mathsf{p}}}(X_1^{n-1}; Y_1^n)$$

$$= \lim_{n\to\infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{\overline{\mathsf{p}}}(X_1^n; Y_1^n) = C(\overline{\mathsf{p}}).$$

The channel space can therefore be reduced to the interval $\mathsf{p} \in [0, \frac{1}{2}]$. Further note that the same symmetry argument holds for not just the rate-maximizing input distribution, but for all input distributions. ∎

The capacity of the Bernoulli state channel is upper bounded by the achievable rate for a genie-aided decoder, i.e., one with the $\mathcal{Z}$ process realization known. Given the realization of the $\mathcal{Z}$ process, the inserted bits and the positions of the deleted bits are known so that the Bernoulli state channel is equivalent to a correlated erasure channel with average erasure rate $\mathbb{P}\{Z_{i-1} = 1, Z_i = 0\} = \mathsf{p}\overline{\mathsf{p}}$. The resulting erasure channel is a correlated channel since, erasures being dependent on $1 \to 0$ transitions in the $\mathcal{Z}$ process, two consecutive bits cannot be erased. Therefore,

$$C(\mathsf{p}) \leq 1 - \mathsf{p}\overline{\mathsf{p}} \triangleq C_{g\varepsilon}(\mathsf{p}) \tag{3}$$

since the capacity of a correlated erasure channel is the same as that of a memoryless erasure channel with the same erasure probability. We call this upper bound, $C_{g\varepsilon}(\mathsf{p})$, the *genie-erasure capacity* of the channel $W(\mathsf{p})$.

Consider

$$\frac{1}{n} I(X_1^n; Y_1^n) = \frac{1}{n} H(Y_1^n) - \frac{1}{n} H(Y_1^n | X_1^n)$$

$$\overset{(a)}{=} \frac{1}{n} H(Y_1^n) - \frac{1}{n} \sum_{i=1}^{n} H(Y_i | X_{i-1}^i)$$

$$= \frac{1}{n} H(Y_1^n) - \frac{h_2(\mathsf{p})}{n} \sum_{i=1}^{n} \mathbb{P}\{X_i \neq X_{i-1}\}, \tag{4}$$

where $h_2(\cdot)$ is the *binary entropy function* [20]. The equality labelled (a) follows from the definition of Y_i in (1) and the fact that $\mathcal{Z} \sim \mathcal{B}(\mathsf{p})$. Since the information-rate-maximizing input distribution is unknown, we will now derive lower bounds to the capacity by making certain assumptions about the statistics of the input process $\mathcal{X}$.

A. I.i.d. input process

We will first assume that the input process $\mathcal{X}$ is an i.i.d. $\mathcal{B}(\alpha)$ process. With this assumption, the maximum achievable information rate, called the *i.i.d. capacity*, denoted $C_{iid}(\alpha, \mathsf{p})$, gives a lower bound to the capacity.

Proposition 2 (Input symmetry): $C_{iid}(\alpha, \mathsf{p}) = C_{iid}(\overline{\alpha}, \mathsf{p})$.

Proof: Let X_1^n and Y_1^n be the input and output respectively of the channel $W(\mathsf{p})$. Define $\hat{X}_1^n = (X_1 \oplus 1, X_2 \oplus 1, \cdots, X_n \oplus 1)$ and $\hat{Y}_1^n = (Y_1 \oplus 1, Y_2 \oplus 1, \cdots, Y_n \oplus 1)$. When

$X_1^n \sim \mathcal{B}(\alpha)$, $\hat{X}_1^n \sim \mathcal{B}(\overline{\alpha})$. Further since $X_1^n \leftrightarrow \hat{X}_1^n$ and $Y_1^n \leftrightarrow \hat{Y}_1^n$ are bijections, we have

$$I_{\mathsf{p}}(X_1^n; Y_1^n) = I(\hat{X}_1^n; \hat{Y}_1^n).$$

Clearly, $\hat{X}_1^n$ and $\hat{Y}_1^n$ also satisfy the relation in (1) and consequently

$$I(\hat{X}_1^n; \hat{Y}_1^n) = I_{\mathsf{p}}(\hat{X}_1^n; \hat{Y}_1^n).$$

From this and the fact that

$$C_{iid}(\alpha, \mathsf{p}) = \lim_{n\to\infty} \frac{1}{n} I_{\mathsf{p}}(X_1^n; Y_1^n)\Big|_{\mathcal{X} \sim \mathcal{B}(\alpha)}$$

we have the desired result. ∎

As a consequence of Propositions 1 and 2, we have

$$C_{iid}(\alpha, \mathsf{p}) = C_{iid}(\alpha, \overline{\mathsf{p}}) = C_{iid}(\overline{\alpha}, \overline{\mathsf{p}}) = C_{iid}(\overline{\alpha}, \mathsf{p}).$$

From Proposition 2 and the fact that $C_{iid}(\alpha, \mathsf{p})$ is concave in α, we immediately have the following.

Corollary 3 (Rate-maximizing i.i.d. distribution):

$$\max_{\alpha \in [0,1]} C_{iid}(\alpha, \mathsf{p}) = C_{iid}(\frac{1}{2}, \mathsf{p}) \ \forall \ \mathsf{p} \in [0, 1].$$

When the binary input is i.i.d. with a uniform distribution (sometimes abbreviated as *i.u.d.*), the corresponding rate is called the *symmetric information rate* (SIR) and is denoted $C_{iud}(\mathsf{p})$. The SIR is of interest because it can be achieved by a random linear coset code [21].

We have from (4)

$$C(\mathsf{p}) \geq C_{iud}(\mathsf{p}) = \mathcal{H}(\mathcal{Y})\Big|_{\mathcal{X} \sim \mathcal{B}(1/2)} - \frac{h_2(\mathsf{p})}{2} \tag{5}$$

We can lower bound the SIR by disregarding the data-dependence of the noise in the channel. This gives a channel equivalent to a BSC with crossover probability $\mathsf{p}/2$ so that

$$C_{iud}(\mathsf{p}) \geq 1 - h_2\left(\frac{\mathsf{p}}{2}\right) \triangleq L_0^{iud}(\mathsf{p}). \tag{6}$$

Further lower bounds can be obtained by conditioning the entropy of the output as follows:

$$C_{iud}(\mathsf{p}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i | Y_1^{i-1}) - \frac{h_2(\mathsf{p})}{2}$$

$$\geq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i | Y_1^{i-1}, X_{i-1}) - \frac{h_2(\mathsf{p})}{2}$$

$$= h_2\left(\frac{1-\mathsf{p}}{2}\right) - \frac{h_2(\mathsf{p})}{2} \triangleq L_1^{iud}(\mathsf{p}), \tag{7}$$

where we have used the fact that Y_i depends only on X_{i-1} and X_i, and given X_{i-1}, Y_i is independent of Y_1^{i-1}. Continuing as above, we can obtain a tighter lower bound

$$C_{iud}(\mathsf{p}) \geq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i | Y_1^{i-1}, X_{i-2}) - \frac{h_2(\mathsf{p})}{2}$$

$$= \left(\frac{1+\mathsf{p}}{2}\right) h_2\left(\frac{1+\mathsf{p}^2}{2(1+\mathsf{p})}\right)$$

$$+ \left(\frac{1-\mathsf{p}}{2}\right) h_2\left(\frac{1-\mathsf{p}}{2}\right) - \frac{h_2(\mathsf{p})}{2} \triangleq L_2^{iud}(\mathsf{p}). \tag{8}$$

A straightforward upper bound for the SIR, implied by (5) is

$$C_{iud}(\mathsf{p}) \le 1 - \frac{h_2(\mathsf{p})}{2} \triangleq U_0^{iud}(\mathsf{p}), \qquad (9)$$

which follows because the entropy rate for a binary process $\mathcal{H}(\mathcal{Y}) \le 1$. Note that this bound is achieved when $\mathcal{Y}$ is the i.i.d. $\mathcal{B}(1/2)$ process. Again, starting from (5), we can obtain upper bounds for the SIR by removing conditioning from the entropy of the output, as shown below:

$$C_{iud}(\mathsf{p}) \le \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i|Y_{i-1}) - \frac{h_2(\mathsf{p})}{2}$$
$$= h_2\left(\frac{1-\mathsf{p}\overline{\mathsf{p}}}{2}\right) - \frac{h_2(\mathsf{p})}{2} \triangleq U_1^{iud}(\mathsf{p}). \qquad (10)$$

As with the lower bounds, we can find a tighter upper bound for the entropy rate $\mathcal{H}(\mathcal{Y})$ as follows

$$C_{iud}(\mathsf{p}) \le \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i|Y_{i-2}^{i-1}) - \frac{h_2(\mathsf{p})}{2}$$
$$= \frac{1-\mathsf{p}\overline{\mathsf{p}}}{2} h_2\left(\frac{1}{2(1-\mathsf{p}\overline{\mathsf{p}})}\right)$$
$$+ \frac{1+\mathsf{p}\overline{\mathsf{p}}}{2} h_2\left(\frac{1}{2(1+\mathsf{p}\overline{\mathsf{p}})}\right) - \frac{h_2(\mathsf{p})}{2} \triangleq U_2^{iud}(\mathsf{p}). \qquad (11)$$

Fig. 6 plots the lower and the upper bounds for SIR discussed above. Note that the tighter lower and upper bounds
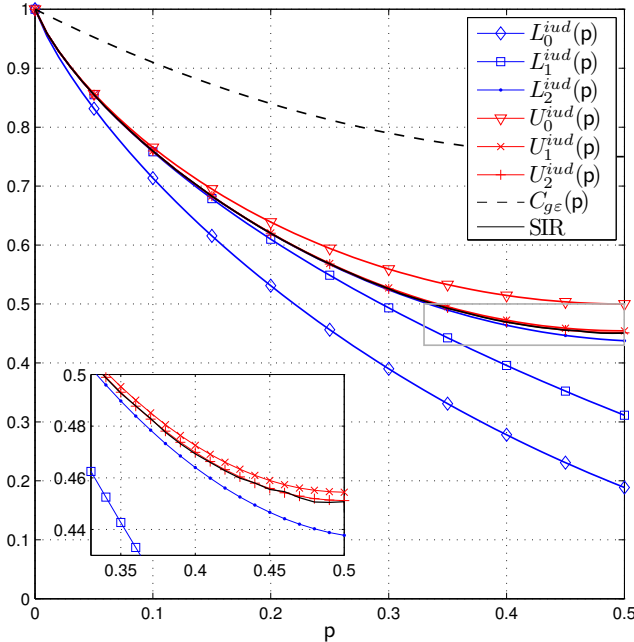

Fig. 6. Lower and upper bounds for SIR given in (6), (7), (8), (9), (10) and (11), and the upper bound for capacity in (3). The SIR from simulations is also shown.

— $L_2^{iud}(\mathsf{p})$ in (8) and $U_2^{iud}(\mathsf{p})$ in (11) — almost coincide for $\mathsf{p} \le 0.3$ (and from symmetry, for $\mathsf{p} \ge 0.7$). In this range, therefore, where the bounds themselves are greater than 0.5, they approximate the SIR fairly accurately. The process $\mathcal{Y}$ is ergodic, as all channel states can be reached within a finite

number of steps at any time with strictly positive probability [22, §5.3], and it converges to a stationary process. As a result, from the Asymptotic Equipartition Property [20, §3.1], we have the entropy rate

$$\mathcal{H}(\mathcal{Y}) = \lim_{n\to\infty} \frac{1}{n} H(Y_1^n) = -\lim_{n\to\infty} \frac{1}{n} \log_2 \mathbb{P}\{Y_1^n\},$$

which can be numerically evaluated through a forward pass of the BCJR algorithm [23], [22, §5.3]. By using long enough sequences X_1^n and Y_1^n in the computation, the SIR can be obtained with an accuracy of $O(\frac{1}{\sqrt{n}})$. This is shown as the "SIR" curve in Fig. 6, from which we conclude that the upper bound $U_2^{iud}(\mathsf{p})$ in (11) is a good approximation for the SIR.

B. First-order Markov Input process

To explore the loss in the achievable rate due to independent uniformly distributed input, we let the source have memory. We consider a first-order binary Markov input process $\mathcal{X}$. Taking a cue from the input symmetry of the Bernoulli state channel, we consider a *symmetric* binary Markov process $\mathcal{X}$ with $\mathbb{P}\{X_i = 1|X_{i-1} = 0\} = \mathbb{P}\{X_i = 0|X_{i-1} = 1\} = \beta$. We denote this by $\mathcal{X} \sim \mathcal{M}_1^{(2)}(\beta)$, meaning that $\mathcal{X}$ is a binary (alphabet of size 2) Markov source with memory 1 and transition parameter β.

Starting from (4), we can arrive at lower and upper bounds for the *Symmetric Markov-1 Rate* (M1R), $C_{M1}(\beta, \mathsf{p})$, which we define as the maximum rate of information transfer when $\mathcal{X} \sim \mathcal{M}_1^{(2)}(\beta)$. The lower bounds analogous to those for the SIR are

$$C_{M1}(\beta, \mathsf{p}) \ge 1 - h_2\left(\frac{\mathsf{p}}{2}\right) \triangleq L_0^{M1}(\mathsf{p}),$$

which is the same as $L_0^{iud}(\mathsf{p})$ in (6),

$$C_{M1}(\beta, \mathsf{p}) = \mathcal{H}(\mathcal{Y})\Big|_{\mathcal{X}\sim\mathcal{M}_1^{(2)}(\beta)} - \beta h_2(\mathsf{p})$$
$$\ge \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i|Y_1^{i-1}, X_{i-1}) - \beta h_2(\mathsf{p})$$
$$= h_2(\beta\overline{\mathsf{p}}) - \beta h_2(\mathsf{p}) \triangleq L_1^{M1}(\beta, \mathsf{p}),$$

and

$$C_{M1}(\beta, \mathsf{p}) \ge \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i|Y_1^{i-1}, X_{i-2}) - \beta h_2(\mathsf{p})$$
$$= (1 - \beta\overline{\mathsf{p}}) h_2\left(\frac{\beta\left(1 - \beta(1-\mathsf{p}^2)\right)}{1 - \beta\overline{\mathsf{p}}}\right)$$
$$+ \beta\overline{\mathsf{p}} h_2(\beta\overline{\mathsf{p}}) - \beta h_2(\mathsf{p}) \triangleq L_2^{M1}(\beta, \mathsf{p}). \qquad (12)$$

The trivial upper bound analogous to $U_0^{iud}(\mathsf{p})$ in (9) is

$$C_{M1}(\beta, \mathsf{p}) \le 1 - \beta h_2(\mathsf{p}) \triangleq U_0^{M1}(\beta, \mathsf{p}).$$

The upper bounds corresponding to U_1^{iud} and U_2^{iud} are

$$C_{M1}(\beta, \mathsf{p}) \le h_2\left(1 - \beta + 2\beta^2\mathsf{p}\overline{\mathsf{p}}\right) - \beta h_2(\mathsf{p}) \triangleq U_1^{M1}(\beta, \mathsf{p})$$

and

$$C_{M1}(\beta, \mathsf{p}) \le \left\{ \beta^2\overline{\beta}(1 - \mathsf{p}\overline{\mathsf{p}}) + \beta^2(3 - \beta)\mathsf{p}\overline{\mathsf{p}} + (1+\beta)\overline{\beta}^2 \right\}$$

$$\times\, h_2\left(\frac{\beta^2\overline{\beta}(1 - \mathsf{p}\overline{\mathsf{p}}) + \beta^3\mathsf{p}\overline{\mathsf{p}} + \beta\overline{\beta}^2}{\beta^2\overline{\beta}(1 - \mathsf{p}\overline{\mathsf{p}}) + \beta^2(3 - \beta)\mathsf{p}\overline{\mathsf{p}} + (1+\beta)\overline{\beta}^2}\right)$$

$$+\, \left\{ 2\beta^2\overline{\beta}(1 - \mathsf{p}\overline{\mathsf{p}}) + \beta^3(1 - 2\mathsf{p}\overline{\mathsf{p}}) + \beta\overline{\beta}^2 \right\}$$

$$\times\, h_2\left(\frac{\beta^2\overline{\beta}(1 - \mathsf{p}\overline{\mathsf{p}}) + \beta^3\mathsf{p}\overline{\mathsf{p}} + \beta\overline{\beta}^2}{2\beta^2\overline{\beta}(1 - \mathsf{p}\overline{\mathsf{p}}) + \beta^3(1 - 2\mathsf{p}\overline{\mathsf{p}}) + \beta\overline{\beta}^2}\right)$$

$$-\, \beta h_2(\mathsf{p}) \overset{\Delta}{=} U_2^{M1}(\beta, \mathsf{p}). \tag{13}$$

Fig. 7 shows the contours of the bounds for C_{M1} in (12) and (13). (Only the tighter bounds, $L_2^{M1}(\beta, \mathsf{p})$ and $U_2^{M1}(\beta, \mathsf{p})$ are shown.) As was the case for i.u.d. input, the tighter bounds
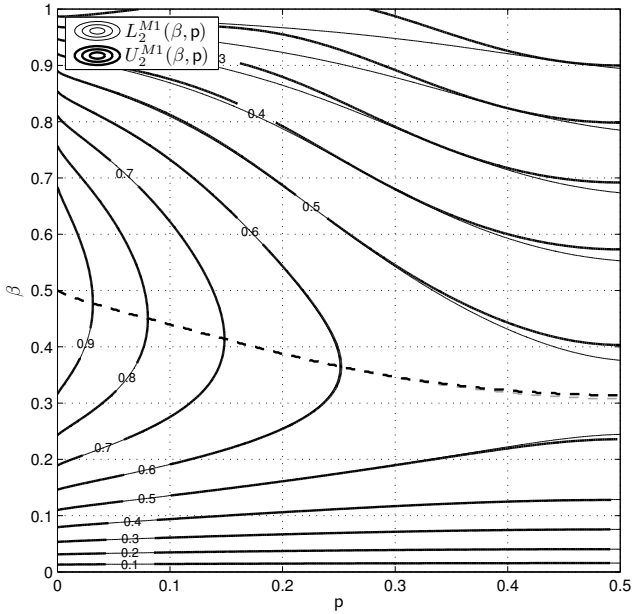

Fig. 7. Contours of lower (thin curves) and upper (thick curves) bounds for M1R given in (12) and (13).

(12) and (13) have almost coinciding contours for a wide range of parameters (β, p). Unlike the case of i.i.d. inputs, the rate-maximizing input parameter $\beta^*(\mathsf{p})$ is not easily obtained. A close estimate can be obtained by maximizing the bounds obtained above. These are shown (dotted lines) in Fig. 7. Since the optimal β values for the tighter lower and upper bounds $L_2^{M1}(\beta, \mathsf{p})$ and $U_2^{M1}(\beta, \mathsf{p})$ almost coincide, we can say that $\beta^*(\mathsf{p})$ is monotonically decreasing in p in the interval $[0, \frac{1}{2}]$ with $\beta^*(0) = \frac{1}{2}$ and $\beta^*(1/2) \approx 0.31$.

Fig. 8 compares the SIR (solid line representing $L_2^{iud}(\mathsf{p})$ in (8), and dashed line representing $U_2^{iud}(\mathsf{p})$ in (11)) and the M1R (solid line for $L_2^{M1}(\beta_{L_2}^*(\mathsf{p}), \mathsf{p})$ in (12), dashed line for $U_2^{M1}(\beta_{U_2}^*(\mathsf{p}), \mathsf{p})$ in (13)) over the range of p values. It is clear that considerable gains in reliable information transfer rate are possible by using an input with memory. In particular, note that whereas there is a range of p values for which the SIR is smaller than 0.5, the M1R is strictly larger than 0.5 $\forall$ p. It is clear that for a sequence of input Markov processes of increasing orders, the achievable rates are non-decreasing. The
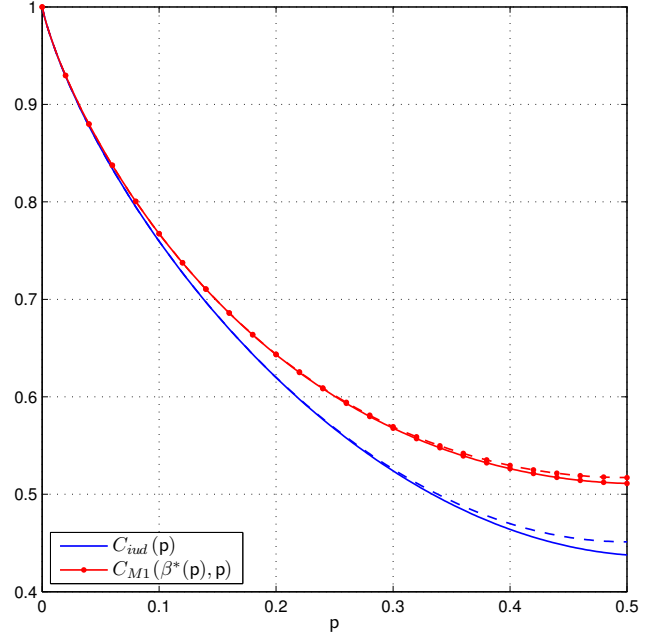

Fig. 8. Comparison between the tighter lower and upper bounds for SIR and M1R.

algorithm suggested in [24], [25] can be employed to optimize the input Markov process to maximize the rate.

V. Binary Markov State Channel

The channel space for the binary Markov state channel defined in Section III-B2 is $(\mathsf{p_i}, \mathsf{p_d}) \in [0, 1]^2$. Note that the channel space is ordered, i.e., the first parameter is the $0 \to 1$ transition (insertion) probability and the second the $1 \to 0$ transition (deletion) probability. As in the case of the Bernoulli state channel, the capacity of the binary Markov state channel, denoted $C(\mathsf{p_i}, \mathsf{p_d})$, is given as

$$C(\mathsf{p_i}, \mathsf{p_d}) = \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{(\mathsf{p_i}, \mathsf{p_d})}(X_1^n; Y_1^n).$$

We shall assume that the channel has converged to the stationary distribution. This means that when $\mathsf{p_i}$ (resp., $\mathsf{p_d}$) is zero, the channel is the noise-free channel (resp., noise-free channel with a delay) and hence $C(0, \mathsf{p_d}) = 1$ (resp., $C(\mathsf{p_i}, 0) = 1$). We first establish the following symmetry property of the binary Markov state channel.

Proposition 4 (Channel symmetry): $C(\mathsf{p_i}, \mathsf{p_d}) = C(\mathsf{p_d}, \mathsf{p_i})$.

Proof: We know that if $Z_1^n = (Z_1, Z_2, \cdots, Z_n)$ is a Markov process, then so is $(Z_n, Z_{n-1}, \cdots, Z_1)$ [26, §16-4]. Furthermore, since the channel is assumed to have converged to the stationary distribution, the conditional distributions $\mathbb{P}\{Z_i|Z_{i-1}\}$ and $\mathbb{P}\{Z_i|Z_{i+1}\}$ are identical. However, note that whereas a transition $Z_{i-1} = 0 \to Z_i = 1$ is an insertion, the transition $Z_{i+1} = 0 \to Z_i = 1$ is a deletion. This implies that $I_{(\mathsf{p_i}, \mathsf{p_d})}(X_i; Y_1^n|X_1^{i-1}) = I_{(\mathsf{p_d}, \mathsf{p_i})}(S_{n-i+1}; Y_1^n|S_{n-i+2}^n)$, where $S_i = X_{i-1}$, for all but a vanishing fraction of indices i, as

$n \to \infty$. Therefore,

$$C(\mathsf{p_i}, \mathsf{p_d}) = \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{(\mathsf{p_i}, \mathsf{p_d})}(X_1^n; Y_1^n)$$

$$= \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} \sum_{i=1}^{n} I_{(\mathsf{p_i}, \mathsf{p_d})}(X_i; Y_1^n | X_1^{i-1})$$

$$= \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} \sum_{i=1}^{n} I_{(\mathsf{p_d}, \mathsf{p_i})}(S_{n-i+1}; Y_1^n | S_{n-i+2}^n)$$

$$= \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{(\mathsf{p_d}, \mathsf{p_i})}(S_1^n; Y_1^n)$$

$$= \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{(\mathsf{p_d}, \mathsf{p_i})}(X_1^{n-1}; Y_1^n)$$

$$= \lim_{n \to \infty} \sup_{\mathbb{P}\{X_1^n\}} \frac{1}{n} I_{(\mathsf{p_d}, \mathsf{p_i})}(X_1^n; Y_1^n) = C(\mathsf{p_d}, \mathsf{p_i}).$$

The channel space can therefore be reduced to the region $\mathsf{p_i} \in [0, 1], \mathsf{p_d} \in [0, \mathsf{p_i}]$. As was the case for the $W(\mathsf{p})$ channel, we have the same symmetry for any fixed input distribution. ∎
As a consequence of Proposition 4, we can assume an unordered pair $\{\mathsf{p_i}, \mathsf{p_d}\}$ parameterizing the channel space.

As for the Bernoulli state channel, we can define the genie-erasure capacity $C_{g\varepsilon}(\mathsf{p_i}, \mathsf{p_d})$ of the binary Markov state channel $W(\mathsf{p_i}, \mathsf{p_d})$. In this case, the resulting channel is a correlated erasure channel with an average erasure probability $\mathbb{P}\{Z_{i-1} = 1, Z_i = 0\} = \frac{\mathsf{p_i}\mathsf{p_d}}{\mathsf{p_i}+\mathsf{p_d}}$, so that

$$C(\mathsf{p_i}, \mathsf{p_d}) \le 1 - \frac{\mathsf{p_i}\mathsf{p_d}}{\mathsf{p_i} + \mathsf{p_d}} \triangleq C_{g\varepsilon}(\mathsf{p_i}, \mathsf{p_d}). \tag{14}$$
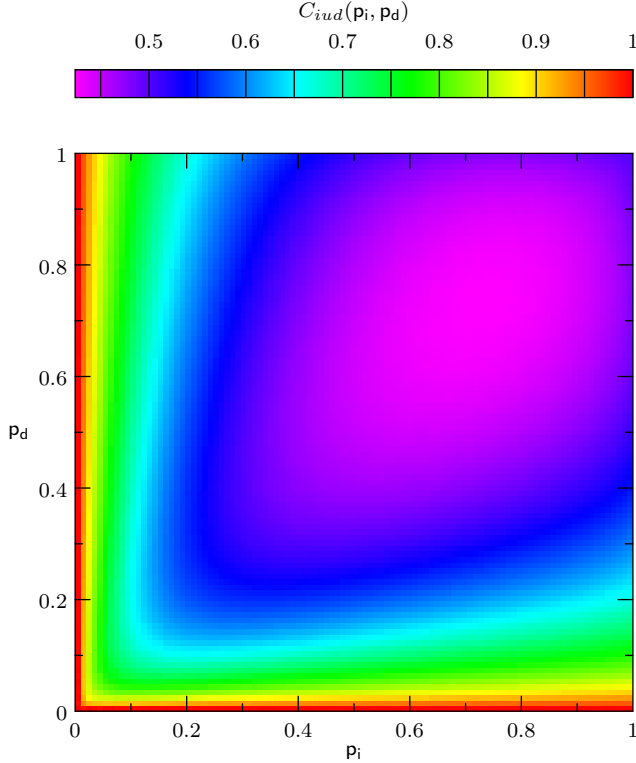
Fig. 9. The SIR $C_{iud}(\mathsf{p_i}, \mathsf{p_d})$ for the $W(\mathsf{p_i}, \mathsf{p_d})$ channel.

As a result of the memory in the $\mathcal{Z}$ process, it is considerably harder than it was for the $W(\mathsf{p})$ channel to obtain closed-form expressions for lower bounds on the capacity of $W(\mathsf{p_i}, \mathsf{p_d})$ by computing information rates for known input distributions. However, the $W(\mathsf{p_i}, \mathsf{p_d})$ channel is still an ergodic channel, and the SIR $C_{iud}(\mathsf{p_{id}}, \mathsf{p_{id}})$ can be obtained numerically. Fig. 9 shows the contours of the SIR for the $W(\mathsf{p_i}, \mathsf{p_d})$ channel. Note that the symmetry proved in Proposition 4 is evident.
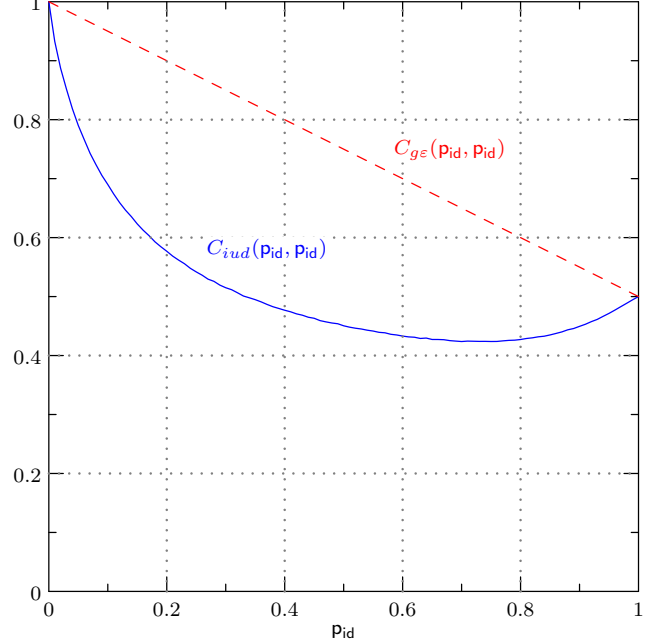

Fig. 10. The SIR $C_{iud}(\mathsf{p_{id}}, \mathsf{p_{id}})$ and the genie-erasure capacity $C_{g\varepsilon}(\mathsf{p_{id}}, \mathsf{p_{id}})$ for the symmetric binary Markov state channel $W(\mathsf{p_{id}}, \mathsf{p_{id}})$.

In Fig. 10, we show the SIR for the case $\mathsf{p_i} = \mathsf{p_d} = \mathsf{p_{id}}$, which we call the *symmetric* binary Markov state channel $W(\mathsf{p_{id}}, \mathsf{p_{id}})$. The values of the SIR when $\mathsf{p_{id}} = 0$ and $\mathsf{p_{id}} = 1$ are easily explained. When $\mathsf{p_{id}} = 0$, the channel is noiseless. When $\mathsf{p_{id}} = 1$, the channel deterministically flips between the identity and the delayed channel so that every odd bit is repeated twice, and every even bit is lost, and the maximum achievable information transfer rate is $\frac{1}{2}$ bit per channel use. Also shown is the genie-erasure capacity in (14), which in this case becomes,

$$C(\mathsf{p_{id}}, \mathsf{p_{id}}) \le C_{g\varepsilon}(\mathsf{p_{id}}, \mathsf{p_{id}}) = 1 - \frac{\mathsf{p_{id}}}{2}.$$

Interestingly enough, when $\mathsf{p_{id}} = 1$, the SIR satisfies $C_{iud}(1, 1) = C_{g\varepsilon}(1, 1)$, so that $C(1, 1) = \frac{1}{2}$. We also include in Fig. 11 the SIR obtained for the channel $W(\mathsf{p_i}, 1)$, as well as the corresponding genie-erasure capacity, $C_{g\varepsilon}(\mathsf{p_i}, 1) = \frac{1}{1+\mathsf{p_i}}$.

VI. K-ARY MARKOV STATE CHANNEL

We now consider a generalization of the binary Markov state channel as described in (2). Here, we allow $Z_i \in \mathbb{Z} = \{0, 1, \cdots, K-1\}$ and let $\mathcal{Z}$ be a first-order Markov process whose transition probabilities satisfy

$$\mathbb{P}\{Z_i = z | Z_{i-1} = z - 1\} = \mathsf{p_i},$$
$$\mathbb{P}\{Z_i = z | Z_{i-1} = z + 1\} = \mathsf{p_d},$$
$$\mathbb{P}\{Z_i = z | Z_{i-1} = z\} = 1 - \mathsf{p_i} - \mathsf{p_d}$$
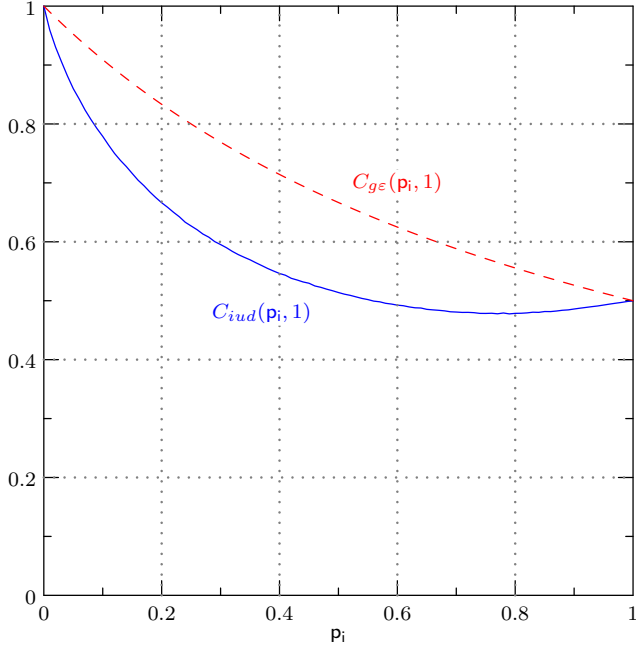
Fig. 11. The SIR $C_{iud}(\mathsf{p_i}, 1)$ and the genie-erasure capacity $C_{g\varepsilon}(\mathsf{p_i}, 1)$ for the channel $W(\mathsf{p_i}, 1)$ with the $\mathcal{Z}$ process satisfying the $(1, \infty)$ constraint.

for $z \in \{1, 2, \cdots, K-2\}$. Further, we have

$$\mathbb{P}\{Z_i = 1 | Z_{i-1} = 0\} = \mathsf{p_i} = 1 - \mathbb{P}\{Z_i = 0 | Z_{i-1} = 0\}$$

and

$$\mathbb{P}\{Z_i = K-2 | Z_{i-1} = K-1\} = \mathsf{p_d}$$
$$= 1 - \mathbb{P}\{Z_i = K-1 | Z_{i-1} = K-1\}.$$

We denote this by $\mathcal{Z} \sim \mathcal{M}_1^{(K)}(\mathsf{p_i}, \mathsf{p_d})$. Note that when $K = 2$, this model gives the binary Markov state channel considered earlier. We will hence be interested in the K-ary Markov state channel where $K > 2$, which we denote by $W(\mathsf{p_i}, \mathsf{p_d}, K)$. Note that this channel now generalizes the binary Markov state channel in the sense that it allows up to $(K-1)$ consecutive insertions or deletions.

We further assume that the parameters $\mathsf{p_i}, \mathsf{p_d}$ are chosen such that the process $\mathcal{Z}$ is aperiodic and irreducible so that $\mathcal{Y}$ is ergodic. The channel space is given by $\mathsf{p_i} \in [0, 1], \mathsf{p_d} \in [0, 1-\mathsf{p_i}]$. The channel symmetry argument of Proposition 4 holds in this case also, so that

$$C(\mathsf{p_i}, \mathsf{p_d}, K) = C(\mathsf{p_d}, \mathsf{p_i}, K)$$

where $C(\mathsf{p_i}, \mathsf{p_d}, K)$ is the capacity of K-ary Markov state channel $W(\mathsf{p_i}, \mathsf{p_d}, K)$. Hence, the channel space can be further reduced to $\mathsf{p_i} \in [0, 1], \mathsf{p_d} \in [0, \min\{\mathsf{p_i}, 1-\mathsf{p_i}\}]$. As was the case for the binary Markov state channel, we have $C(0, \mathsf{p_d}, K) = C(\mathsf{p_i}, 0, K) = 1 \ \forall \ \mathsf{p_i}, \mathsf{p_d} \in [0, 1]$.

The genie-erasure capacity of the $W(\mathsf{p_i}, \mathsf{p_d}, K)$ channel is given by the capacity of a correlated erasure channel with an

average erasure rate

$$\mathbb{P}\{Z_i = Z_{i-1} - 1\} = \sum_{z=1}^{K-1} \mathbb{P}\{Z_{i-1} = z, Z_i = z-1\}$$

$$= \sum_{z=1}^{K-1} \mathbb{P}\{Z_{i-1} = z\} \mathbb{P}\{Z_i = z-1 | Z_{i-1} = z\} = \sum_{z=1}^{K-1} \pi_z \mathsf{p_d}$$

$$= \mathsf{p_d} \left(1 - \left(\frac{\mathsf{p_d}}{\mathsf{p_i}} \right)^{K-1} \left(\frac{1 - \left(\frac{\mathsf{p_d}}{\mathsf{p_i}} \right)}{1 - \left(\frac{\mathsf{p_d}}{\mathsf{p_i}} \right)^K} \right) \right) = \mathsf{p_\varepsilon}$$

where $\pi_z = \left(\frac{\mathsf{p_d}}{\mathsf{p_i}} \right)^{K-1-z} \left(\frac{1 - \left(\frac{\mathsf{p_d}}{\mathsf{p_i}} \right)}{1 - \left(\frac{\mathsf{p_d}}{\mathsf{p_i}} \right)^K} \right)$ is the steady state probability of $Z_i = z$, so that

$$C(\mathsf{p_i}, \mathsf{p_d}, K) \leq 1 - \mathsf{p_\varepsilon} \overset{\Delta}{=} C_{g\varepsilon}(\mathsf{p_i}, \mathsf{p_d}, K).$$

For the *symmetric* K-ary Markov state channel, we have

$$C(\mathsf{p_{id}}, \mathsf{p_{id}}, K) \leq 1 - \mathsf{p_{id}} \frac{K-1}{K} = C_{g\varepsilon}(\mathsf{p_{id}}, \mathsf{p_{id}}, K)$$

for $\mathsf{p_{id}} \in [0, \frac{1}{2}]$ because in this case, $\pi_z = \frac{1}{K} \ \forall \ z \in \{0, 1, \cdots, K-1\}$. Note that $C_{g\varepsilon}(\mathsf{p_{id}}, \mathsf{p_{id}}, K)$ reduces to the capacity of a BEC with erasure probability $\mathsf{p_{id}}$, $C_{BEC}(\mathsf{p_{id}}) = 1 - \mathsf{p_{id}}$, when $K \to \infty$.

The SIR of the $W(\mathsf{p_i}, \mathsf{p_d}, K)$ channel can be obtained numerically as in the case of the $W(\mathsf{p_i}, \mathsf{p_d})$ channel. However, the computational complexity of the BCJR algorithm [27] used to estimate the SIR increases roughly exponentially in the size K of the alphabet of the process $\mathcal{Z}$, and hence the evaluation of the SIR for $K > 2$ may require considerable computing resources. We carried out the calculation for the case when $K = 3$, and the estimated SIR is shown in Fig. 12.
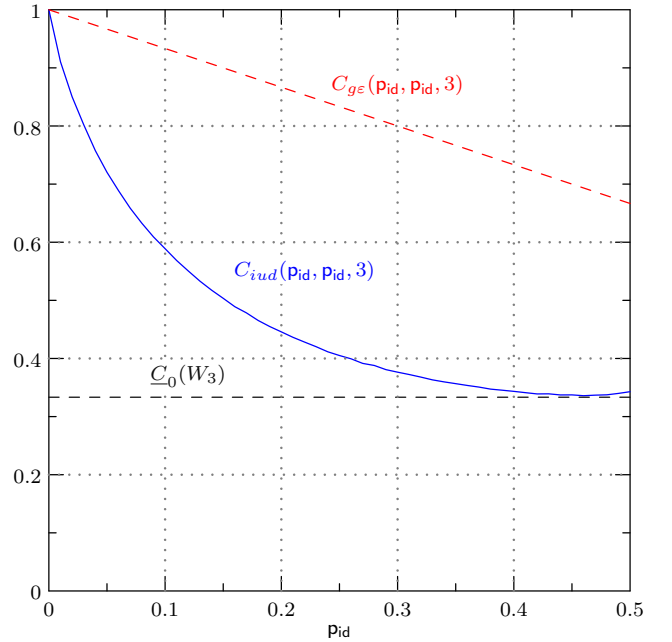

Fig. 12. The SIR $C_{iud}(\mathsf{p_{id}}, \mathsf{p_{id}}, 3)$ for the symmetric 3-ary Markov state channel, along with the genie-erasure capacity $C_{g\varepsilon}(\mathsf{p_{id}}, \mathsf{p_{id}}, 3)$. Also shown is the lower bound on the zero-error capacity $\underline{C}_0(W_3)$ from Corollary 6.

VII. Zero-error capacity

The *zero-error capacity* C_0 of a discrete-memoryless channel was introduced by Shannon [28], from which it is readily seen that a noisy discrete-output binary-input memoryless channel has a zero-error capacity of 0, i.e., no information can be transmitted over such a channel with zero error. However, we will now show that the zero-error capacity of the noisy discrete-output binary-input channel in (1) is strictly positive. We will denote the generic channel in (1) for all binary $\mathcal{Z}$ processes by W.

Proposition 5: $C_0(W) = \frac{1}{2}$.

Proof: Consider the 2-repetition code. The channel input process $\mathcal{X}$ and the message process $\mathcal{M}$ satisfy the relationship $X_{2i-1} = M_i, X_{2i} = M_i$ for $i = 1, 2, \cdots$. Since $X_{2i-1} = X_{2i}$, $Y_{2i} = X_{2i} = M_i$ so that discarding the Y_{2i-1}s gives us $\mathcal{M}$ exactly, thereby achieving zero error at a rate $\frac{1}{2}$. Thus, $C_0(W) \geq \frac{1}{2}$.

It is obvious that the capacity of a channel is an upper bound on the zero-error capacity. From (14), we can see that $C(1,1) \leq \frac{1}{2}$, and the coding scheme above achieves a rate 0.5 for any realization of $\mathcal{Z}$. Thus, $C(1,1) = \frac{1}{2}$. Therefore, as long as the channel admits the infinitely long alternating sequence $\mathcal{Z} = 010101\cdots$, we have $C_0(W) = \frac{1}{2}$.

Note that both the Bernoulli state channel $W(\mathsf{p})$ and the binary Markov state channel $W(\mathsf{p_i}, \mathsf{p_d})$ can generate the infinite sequence $\mathcal{Z} = 010101\cdots$, albeit with vanishing probability. It is also worth noting here that this infinite sequence satisfies the $(1, \infty)$ constraint (See Section III-C). ∎

Thus, unlike binary-input discrete memoryless channels, the binary-input discrete channel in (1) allows a non-zero information rate even under the severe requirement of zero-error. From this result and Fig. 8, it is clear that even under the milder condition of asymptotically vanishing error probability, random linear coset coding achieves a lower rate than the 2-repetition code, which guarantees zero errors, for the Bernoulli state channel over a range of p values. Referring to Fig. 9, the same can be said for the binary Markov state channels. However, by using a first-order Markov input (cf. Fig. 8), higher rates than that of the zero-error achieving scheme are achievable for all Bernoulli state channels, although only with asymptotically vanishing error probability.

The zero-error capacity of the $W(\mathsf{p_i}, \mathsf{p_d}, K)$ channel, denoted by $C_0(W_K)$, satisfies the following bounds.

Corollary 6: $\frac{1}{K} \leq C_0(W_K) \leq \frac{K+1}{2K}$.

Proof: Repeating every bit K times achieves zero error, since every K^{th} bit is always correct for any realization of the $\mathcal{Z}$ process. Hence, $C_0(W_K) \geq \frac{1}{K} = \underline{C}_0(W_K)$.

The smallest upper bound for the capacity is also an upper bound for the zero-error capacity. Thus,

$$
\begin{aligned}
C_0(W_K) &\leq \min_{\mathsf{p_i}, \mathsf{p_d}} C(\mathsf{p_i}, \mathsf{p_d}, K) \\
&\leq \min_{\mathsf{p_i}, \mathsf{p_d}} C_{g\varepsilon}(\mathsf{p_i}, \mathsf{p_d}, K) \\
&= C_{g\varepsilon}\left(\frac{1}{2}, \frac{1}{2}, K\right) = \frac{K+1}{2K}.
\end{aligned}
$$

Observe that when $K = 2$, the minimum genie-erasure capacity occurs at $\mathsf{p_i} = \mathsf{p_d} = 1$ and at this point the upper bound $C_{g\varepsilon}(1, 1, 2)$ is same as the lower bound $\underline{C}_0(W_2)$ which was used in the proof of Proposition 5. ∎

Note that the bounds proposed above are loose asymptotically, i.e., $\frac{1}{K} \to 0$ and $\frac{K+1}{2K} \to \frac{1}{2}$ as $K \to \infty$.

VIII. Conclusions

We proposed a new write channel model for bit-patterned media recording that reflects the data dependence of write synchronization errors. This model generates both substitution-like errors and insertion-deletion errors whose statistics are determined by an underlying channel state process. For Bernoulli and Markov channel state models, we studied information-theoretic limits imposed by the channel, computing bounds and numerical estimates for the maximum achievable information rate under different assumptions on the channel input statistics. We then generalized the Markov channel state model to allow a channel state space of size $K > 2$ and computed the SIR numerically for the case $K = 3$. Finally, we showed that the rate-$\frac{1}{2}$ 2-repetition code achieves the zero-error capacity over the new write channel when the channel state space is binary. Bounds on the zero-error capacity of the general K-ary Markov channel state model were also presented. In future work, we plan to combine the new write channel model in cascade with the partial-response readback channel model and investigate the achievable rates, as was done for the i.i.d. insertion-deletion channel in [16]. Furthermore, the channel model considered here can be readily represented using a factor graph and hence the construction of sparse graph-based codes with iterative decoding techniques seems possible.

References

[1] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "Data-dependent write channel model for magnetic recording," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 13-18, 2010, pp. 958–962.

[2] H. Kobayashi, "A survey of coding schemes for transmission or recording of digital data," *IEEE Trans. on Comm. Techn.*, vol. 19, no. 6, pp. 1087–1100, Dec. 1971.

[3] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Trans. Commun.*, vol. 23, no. 9, pp. 921–934, Sep. 1975.

[4] J. Hu, T. Duman, E. Kurtas, and M. Erden, "Bit-patterned media with written-in errors: modeling, detection, and theoretical limits," *IEEE Trans. Magn.*, vol. 43, no. 8, pp. 3517–3524, Aug. 2007.

[5] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, "LDPC codes for the cascaded BSC-BAWGN channel," in *Proc. 47th Annual Allerton Conf. on Communication, Control and Computing*, Sep. 30 - Oct. 2, 2009, pp. 620–627.

[6] R. White, R. Newt, and R. Pease, "Patterned media: a viable route to 50 Gbit/in² and up for magnetic recording?" *IEEE Trans. Magn.*, vol. 33, no. 1, pp. 990–995, Jan 1997.

[7] B. Livshitz, A. Inomata, H. Bertram, and V. Lomakin, "Semi-analytical approach for analysis of BER in conventional and staggered bit patterned media," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3519–3522, Oct. 2009.

[8] M. Mitzenmacher, "Capacity bounds for sticky channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 72–77, Jan. 2008.

[9] S. Diggavi and M. Grossglauser, "On transmission over deletion channels," in *Proc. 39th Annual Allerton Conf. on Communication, Control and Computing*, Oct. 3-5, 2001, pp. 573–582.

[10] ——, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, Mar. 2006.

[11] E. Drinea and M. Mitzenmacher, "On lower bounds for the capacity of deletion channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4648–4657, Oct. 2006.

[12] M. Mitzenmacher and E. Drinea, "A simple lower bound for the capacity of the deletion channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4657–4660, Oct. 2006.

[13] S. Diggavi, M. Mitzenmacher, and H. D. Pfister, "Capacity upper bounds for the deletion channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 24-29, 2007, pp. 1716–1720.

[14] D. Fertonani and T. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2753 – 2765, Jun. 2010.

[15] A. Kirsch and E. Drinea, "Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 86 –102, Jan. 2010.

[16] J. Hu, T. Duman, M. Erden, and A. Kavcic, "Achievable information rates for channels with insertions, deletions, and intersymbol interference with i.i.d. inputs," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1102 –1111, Apr. 2010.

[17] R. Wood, M. Williams, A. Kavcic, and J. Miles, "The feasibility of magnetic recording at 10 Terabits per square inch on conventional media," *IEEE Trans. Magn.*, vol. 45, no. 2, pp. 917–923, Feb. 2009.

[18] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Shannon Foundation Publishers, The Netherlands, 1999.

[19] A. Mazumdar, A. Barg, and N. Kashyap, "Coding for high-density magnetic recording," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 13-18, 2010, pp. 978–982.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. New York: John Wiley and Sons, 2006.

[21] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.

[22] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[23] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[24] A. Kavcic, "On the capacity of Markov sources over noisy channels," in *Proc. IEEE Globecom*, vol. 5, San Antonio, TX, USA, Nov. 25-29, 2001, pp. 2997–3001.

[25] P. Vontobel, A. Kavcic, D. Arnold, and H.-A. Loeliger, "A generalization of the Blahut-Arimoto algorithm to finite-state channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1887 –1918, May 2008.

[26] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill Inc., 1991.

[27] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, "Determining and approaching achievable rates of binary intersymbol interference channels using multistage decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1416–1429, Apr. 2007.

[28] C. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.